**EDA + Visualization II**

Zoom Recording:

<https://berkeley.zoom.us/rec/share/yGOq9WiCKnnfT8FrbuGnMmed6mDBHwr2Bv4A3JF2iQAXpsT-AhTFDuaJWY07EllX.5ZuS1-TtA4xsT0O3>

This week mainly serves as an extension to last week, where we continue applying similar techniques to a new dataset, and try to draw conclusions from the data provided. See if you can try out new ways to visualize the data, and if there are any similarities or differences compared to what you’ve found in the heart disease dataset.

Here are some possible areas you might want to explore more on your own.

* Exploring the relationship between other categorical variables such as smoking, diabetes, etc.
* You may also wish to investigate further into serum creatinine and ejection fraction, which both seem to correlate closely to heart failure. See if you can find any relationships between these and other variables as well!
* You may also wish to spend some time going back to the heart disease dataset if there were anything you might still be interested in exploring more on.
* Anything else you can come up with on your own.

At this point, you should feel comfortable with the general idea behind both datasets, and begin to develop a better idea on how different features relate to each other in both of these, as well as having an understanding of how you might go about approaching different datasets in the future.

[Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone | BMC Medical Informatics and Decision Making](https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-020-1023-5)